

Link Characteristics that Promote Clicks

STATCOM

Link Characteristics that Promote Clicks: Executive Summary

Questions of Interest

- Which link characteristics (bolded, font size, font color, location within newsletter) lead to more clicks?
- Which topics mentioned in the links (e.g. animals, art, family, hunger) tend to attract clicks?

Statistical Analysis

- Bar plots and scatter plots to visually display click response based on link characteristics
- Statistical model to examine how factors interact and support the trends shown in the plots
- Qualitative analysis on the words associated with a link

Takeaways

- The location of the link in the newsletter affects how often it is clicked. Specifically, the further down the link is in the newsletter, the less likely it is to be clicked on. Therefore, we suggest placing the most important or time-sensitive links at the top of the newsletter.
- Links that are bolded get slightly more clicks than links that do not have bolded text.
- Links that are orange [Mandarin Orange (146, 46, 33), Tangerine (242, 136, 0) or (238, 135, 2), Orange Peel (255, 151, 9)], gray [Charcoal (67, 67, 67)], or blue [Denim (17, 85, 204), Danube (85, 142, 190)] colored get more clicks than other colors.
- Having the link address appear in the newsletter multiple times, whether as text or image links, increases the number of clicks. However, this effect is not significant, according to the model.
- People tend to click on text links rather than image links.
- Opportunities with baby chicks and sharing one's skills are significantly popular.

Link Characteristics that Promote Clicks: Report

In this report, we investigate the characteristics of links that make them more likely to be clicked on.

The rest of the report is organized as follows. First, we give a brief description of how the data was obtained and a synopsis of the assumptions we made to analyze the click data. Then, we introduce the features used in the model and analyze how click rates were affected by these features separately. Finally, we fit a statistical model to the data and interpret the results.

Data Description

The data comes from a few sources: the CSV files generated from iContact and the plain-text and HTML source code of each of the newsletters. From the CSV files, we determine the unique number of times a link was clicked on. We define a unique click to be a unique combination of subscriber ID, newsletter date, and link; in other words, if a subscriber clicked on the same link from the same newsletter, we do not count that click.

Additionally, we only know the click count for each link address, not the actual link, within a newsletter. Thus, if there are multiple links with the same address in a newsletter, we don't know how many clicks each of those separate links received. To alleviate this issue of duplicate addresses, we assume that the first text link with a given address received all the clicks associated with that address. We account for whether the address is duplicated in our model.

We identify the time of day the newsletter was sent and whether it was before or after the COVID-19 pandemic was declared (03/20/20) from the CSV files. The plain-text files are used to get the text associated with a link and an approximation of how far down the newsletter the link is, e.g. a link that is about half-way down the newsletter would be assigned 50%. Finally, we obtain style characteristics and whether the link was an image or had an image associated with it from the HTML source code.

Below is a summary of the features we created for the text links:

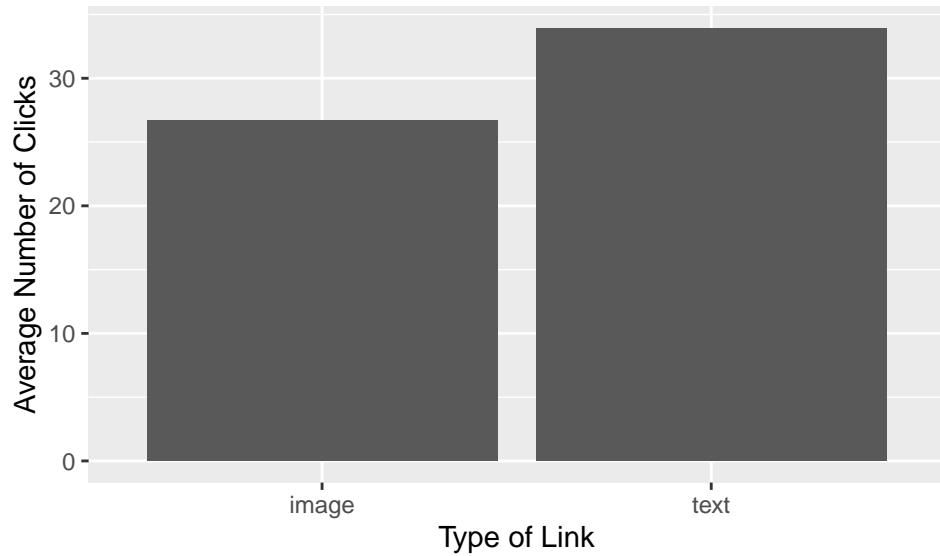
1. Bolded: whether the link text is bolded.
2. Font Size: ranges from 10-48 point.
3. Font Color: 26 possible colors. See Appendix for RGB values.
4. Image Associated: indicator for whether there is an image within the newsletter with the same link address.
5. Hour: hour of when the newsletter was sent
6. COVID: indicator for whether the COVID-19 pandemic was underway
7. Location within document: cumulative percentage of the document prior to a link
8. Duplicate: indicator for whether the link address appears more than once in a newsletter

Data Exploration

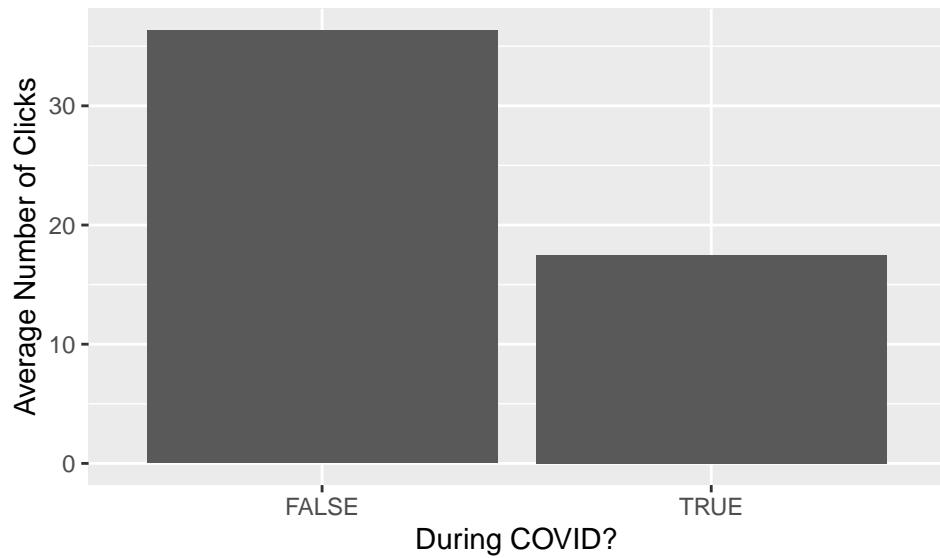
Before fitting any models to the data, we explore how the number of clicks a link address receives depends on the variables mentioned above. It is important to note that in doing this, we do not control for how many times a link address was used within the same newsletter. For each of the categorical variables, we graph the category and the average number of times a link of that category was clicked below.

All Links

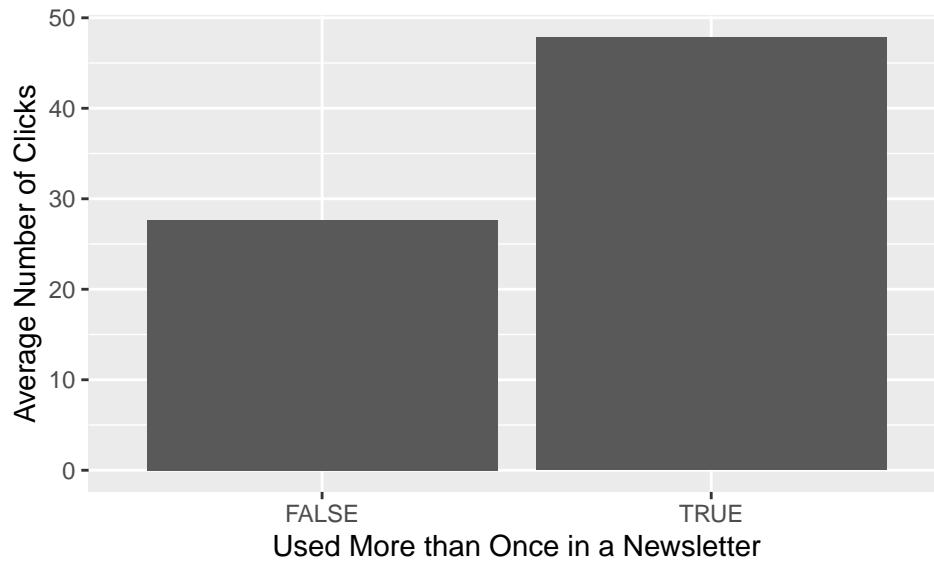
The following barplots pertain to all links, whether they are text or images.



Text links are clicked on more often than pictures. However, because our model only examines text links, we cannot verify this trend with the model.



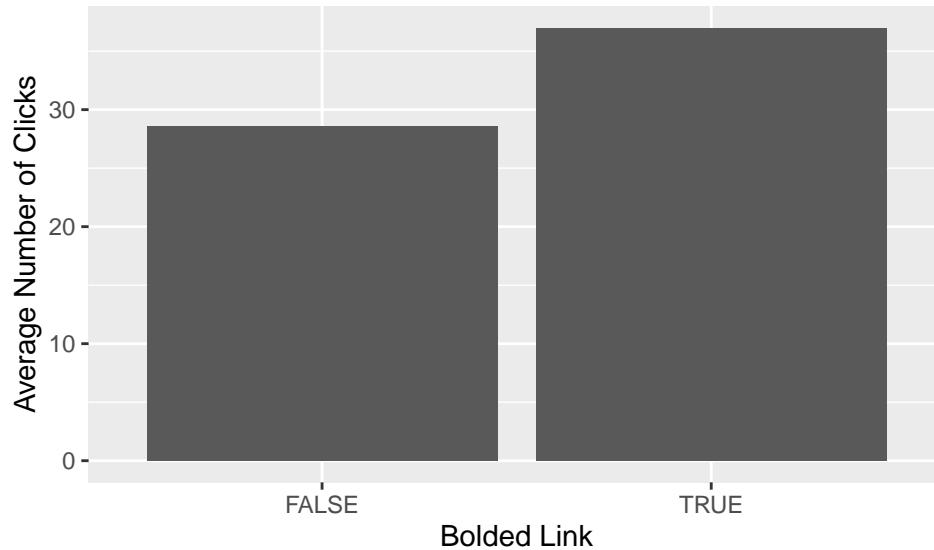
Based on the bar plot above, it seems COVID impacted how much subscribers choose to interact with the newsletters. This is not all that surprising given the challenges everyone was facing during the pandemic.



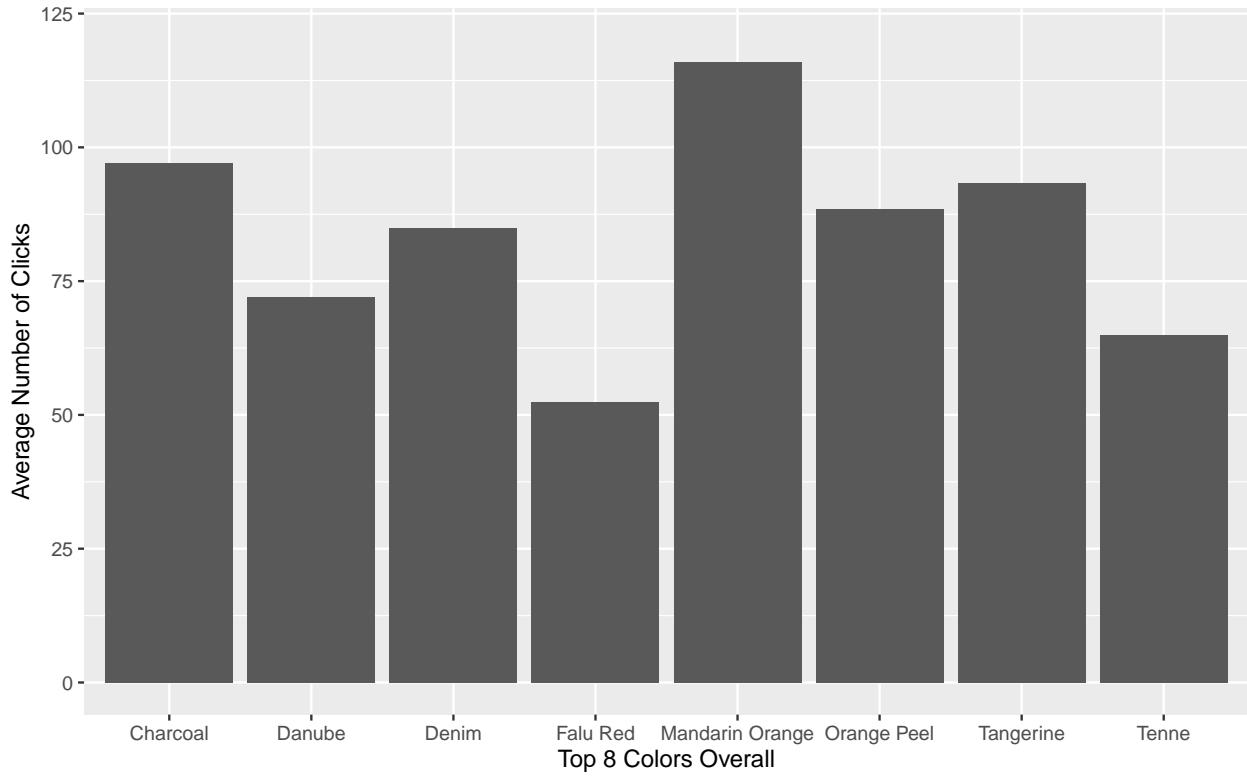
The above bar plot indicates that addresses that appear multiple times in a newsletter tend to be clicked on more often, as expected. Even so, this effect is not significant, according to the below model.

Text Links

The following barplots pertain to only text links.

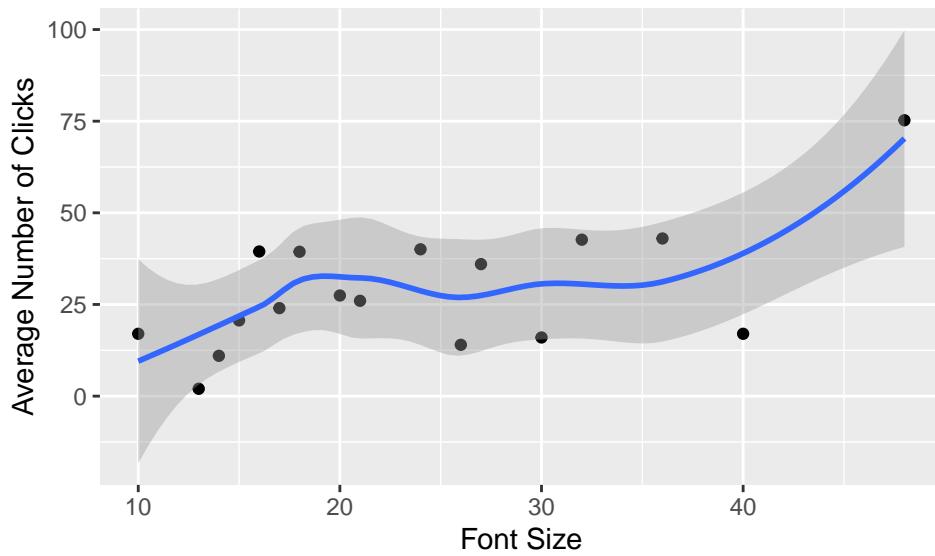


Based on the bar plot above, it appears the bolding the text associated with the link also increases the chance that someone clicks on it.

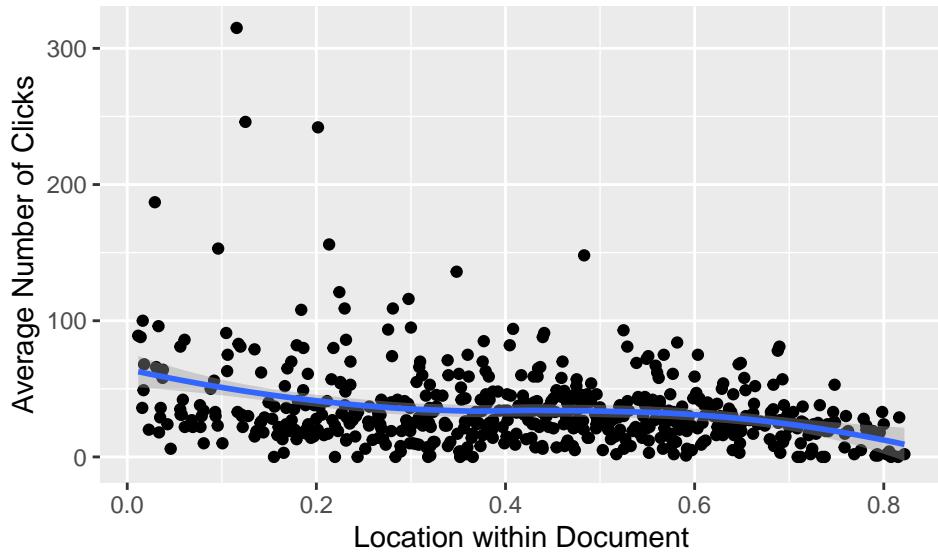


This bar plot focuses on the top eight most clicked-on text color choices, on average. See Appendix for RGB values.

Any color of orange seems to grab people's attention! Mandarin Orange only appeared in the newsletter promoting the Remote Volunteer Project: DIY Family Essentials Kits opportunity so it is tempting to think the large number of clicks this color received may have more to do with the highly-relatable project. However, this project was advertised in four different newsletters using links colored as cinnabar and falu red (both are different tints of red) and these links were not clicked on as often. While there are more factors at play than just the link color, the fact that the links advertising the same opportunity in red got fewer clicks suggests that a text color of orange is more impactful.



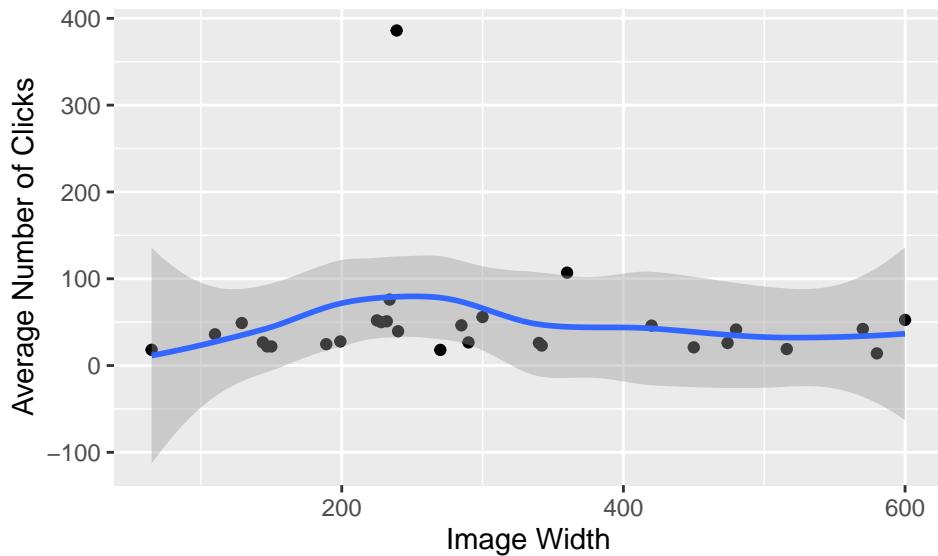
Larger links seem to be clicked on more. However, according to the below model, this effect is not significant.



The further down the newsletter link is, the less likely it is to be clicked on.

Image Links

The following barplots pertain to only image links.



The dimensions of the image is determined by the image width—the image height is adjusted automatically. According to the scatter plot, the image width doesn't appear to affect the click rate.

Model Fitting

We fit a zero-inflated beta regression model to uncover trends regarding the text links. This model does not include image links.

The outcome variable of interest, the number of clicks, was standardized by dividing the click count by the number of subscribers the newsletter containing the link was sent to. The beta regression allows us to model proportion data (data that's bounded between zero and one, non-inclusive). The “zero-inflated” in the name

refers to extending the beta regression to include observations with a value of zero (many links received no clicks). The zero-inflated beta regression fits three parameters: mu, sigma, and nu. The mu variable corresponds to the mean of the click proportion (relative to the number of subscribers) and is modeled in a similar manner to simple linear regression.

The variables in our model are the following: doc_prop, bolded, color, font_size, hour, covid_ind, imag_assoc, and dup. “doc_prop” is the proportion down the document a link is; in other words, a link that is about halfway down a newsletter will be about 50%. “bolded” indicates whether a link was bolded. “color” is the color of the link as named by <https://www.color-blindness.com/color-name-hue/>.

Below we give a histogram of click proportion and the fitted model parameters for mu. From Table 1 below, we see that the link’s location in the document, whether the link is bolded, and the color of the link make a statistically significant difference on whether the link is clicked or not. Additionally, we see that the top eight colors shown above are also statistically significant, except for Tenne. Interestingly, the indicator for whether the link address is duplicated or not, nor the indicator for whether the link address has an image associated with it, are not significant in the model.

Finally, the mu coefficients given in the table below are, unfortunately, uninterpretable in their raw form. Luckily, a transformation of these coefficients gives the odds ratio of each variable. For the variables that are statistically significant at or below the 0.01 level, we give the odds ratios in Table 2. As an example of how odds ratios are interpreted, the doc_prop odds ratio of 0.46 means that when the location of the link in the document increases by one percentage point, the odds of it being clicked is 0.46 times the odds of it being clicked in the original position. Overall, an odds ratio greater than one indicates a positive association and an odds ratio less than one indicates a negative association; note that these agree with the signs of the coefficients in Table 1.

Histogram of Click Proportions

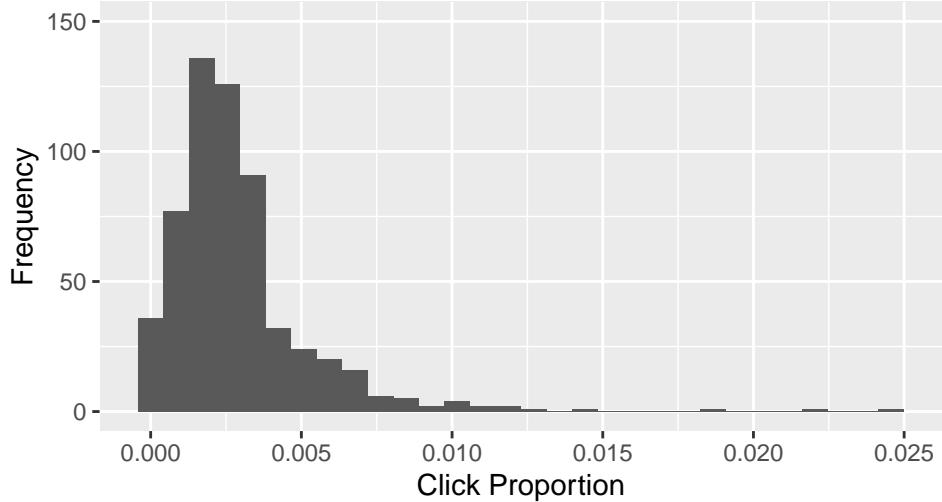


Table 1: Mu Coefficients

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.96	0.26	-23.18	< 2e-16 ***
doc_prop	-0.78	0.14	-5.56	0.00 ***
boldedTRUE	0.22	0.06	3.50	0.00 ***
color_nameBlack	0.61	0.36	1.71	0.09 .
color_nameBlack Pearl	-0.09	0.44	-0.21	0.84
color_nameCharcoal	1.06	0.31	3.39	0.00 ***
color_nameChocolate	0.47	0.24	1.93	0.05 .
color_nameCinnabar	0.07	0.33	0.23	0.82
color_nameCitron	0.13	0.30	0.45	0.66
color_nameDanube	1.62	0.34	4.72	0.00 ***
color_nameDenim	1.49	0.31	4.84	0.00 ***
color_nameDim Gray	-0.23	0.73	-0.32	0.75
color_nameEastern Blue	0.28	0.25	1.09	0.28
color_nameEclipse	0.37	0.24	1.58	0.12
color_nameFalu Red	0.64	0.24	2.70	0.01 **
color_nameGamboge	0.15	0.53	0.29	0.77
color_nameGrey	0.18	0.24	0.77	0.44
color_nameMandarin Orange	1.35	0.38	3.52	0.00 ***
color_nameMariner	0.26	0.59	0.43	0.67
color_nameNero	0.18	0.68	0.27	0.79
color_nameOrange Peel	1.04	0.34	3.08	0.00 **
color_nameSlate Blue	0.45	0.61	0.75	0.46
color_nameTangerine	0.94	0.28	3.42	0.00 ***
color_nameTeal	0.44	0.24	1.82	0.07 .
color_nameTenne	0.85	0.44	1.92	0.06 .
color_nameTyrian Purple	0.28	0.55	0.51	0.61
color_nameWhite	0.42	0.24	1.72	0.09 .
font_size	-0.01	0.01	-1.03	0.31
hour	0.01	0.01	1.29	0.20
covid_indTRUE	-0.70	0.08	-9.17	< 2e-16 ***
image_assocTRUE	0.03	0.10	0.32	0.75
dupTRUE	0.12	0.09	1.30	0.19

Table 2: Odds Ratio

	Estimate	Odds Ratio
(Intercept)	-5.96	0.003
doc_prop	-0.78	0.460
boldedTRUE	0.22	1.240
color_nameCharcoal	1.06	2.884
color_nameDanube	1.62	5.061
color_nameDenim	1.49	4.445
color_nameFalu Red	0.64	1.896
color_nameMandarin Orange	1.35	3.845
color_nameOrange Peel	1.04	2.817
color_nameTangerine	0.94	2.571
covid_indTRUE	-0.70	0.495

Qualitative Text Analysis

Finally, we explored what words encouraged subscribers to click on a link by creating word clouds. The word clouds are composed of the capitalized words that were contained in each link. By focusing on capitalized words, we aim to ignore most of the unimportant filler words. The size of the words corresponds to the proportion of unique clicks relative to the number of total clicks a newsletter obtained. Note that this is slightly different than the proportion of clicks defined above for the zero-inflated beta model.

The first word cloud below is for any links that were not social media for Activate Good.



The second word cloud below is for links that were for opportunities as defined by the link containing “opportunity” in the address; these links correspond to volunteer opportunities for subscribers.



As we can see, opportunities with baby chicks and sharing one's skills are significantly popular compared to other opportunities!

Appendix (Color RGB Values and Opportunity Link Click Frequency Table)

Here we provide the RGB values for all of the colors that were present in the news letter links. The names were assigned according to <https://www.color-blindness.com/color-name-hue/>.

Color Names
Bahia (179,183,27)
Black (0,0,0)
Black (10,10,10)
Black Pearl (29,33,41)
Charcoal (67,67,67)
Chocolate (228,104,16)
Chocolate (233,93,20)
Cinnabar (231,93,38)
Citron (152,154,38)
Danube (85,142,190)
Denim (17,85,204)
Dim Gray (97,97,97)
Eastern Blue (0,136,168)
Eclipse (55,55,55)
Eclipse (57,57,57)
Falu Red (148,45,27)
Gamboge (228,134,9)
Grey (127,127,127)
Mandarin Orange (146,46,33)
Mariner (56,88,152)
Mortar (85,85,85)
Mortar (89,89,89)
Nero (34,34,34)
Orange Peel (255,150,0)
Orange Peel (255,151,9)
Slate Blue (102,94,208)
Tangerine (238,135,2)
Tangerine (242,124,0)
Tangerine (242,136,0)
Tangerine (248,118,0)
Teal (0,109,131)
Tenne (206,86,0)
Tyrian Purple (13,0,0)
White (255,255,255)